Covariate balancing strategy for single and multiple exposures with interaction

Yan-ni Jhan[1], *Thai Son* Dinh[2], and *Ie-bin* Lian[1*]

[1]Data Research Center / Institute of Statistics and Information Science, National Changhua University of Education, Taiwan.
[2]Lac Hong University, Vietnam.

Abstract. Balancing the distribution of covariates (Z) among exposure levels is a crucial step for establishing causality between the exposure and the outcome in observational studies. Standard approaches utilizing propensity score typically focus on a single exposure, yet it is not uncommon for the exposure to interact with other variables on the outcome. Ignoring such interactions and applying standard balancing procedures solely on a single exposure can lead to significant bias. For instance, consider the Georgia Capital Charging and Sentencing Study, which sought to examine whether the race of the defendant and the race of the victim influenced the severity or length of the sentence (Y). In such a study, there are two exposures of interest on the outcome with significant interaction. Analysing each exposure separately may produce biased results. Base on the simulation results we suggest to use covariate-partition strategy for single-exposure scenario and all-covariate strategy for multiple-exposure scenario.

1 Introduction

The balancing score (denoted as e) and entropy balance (EB) weighting are popular and efficient approaches for controlling confounding effects in observational causal analysis with many covariates [1]. For response Y, we assume X to be the exposure variable of interest and $\boldsymbol{Z}$ to be the p available covariates of sample n. For binary X (e.g., exposure vs. non-exposure), the balancing score estimated by fitting the logistic regression of X on $\mathbf{Z}$ is known as the propensity score (PS) [2-4]. The procedure of applying PS usually contains two steps:

Step 1. Estimating PS as the prediction of X by a function of $\boldsymbol{Z}$ (denoted as $PS(Z)$).

Step 2. Inferencing the $X-Y$ association with one of the following balancing methods by $PS(Z)$:

- 2.1 Regression adjustment (PS-reg): to regress Y on X and $PS(Z)$.
- 2.2 Stratification on $PS(Z)$.
- 2.3 Matching by $PS(Z)$.
- 2.4 Inverse probability of treatment weighting (IPTW), $\frac{X}{PS(Z)} + \frac{1-X}{1-PS(Z)}$.
- 2.5 Doubly robust (DR) estimation.

* Corresponding author: maiblian@cc.ncue.edu.tw

DR estimation is an approach that requires regression models to be specified for the outcome and the exposure as a function of covariates $\boldsymbol{Z}$ [5-8]. It combines a form of regression with an estimation of PS (step 1), and inference of the causal effect of an exposure on an outcome (step 2) using the following estimator:

$$\frac{1}{n}\sum\left[\frac{X_j(Y_{j,}-\hat{\mu}_1(\boldsymbol{Z}_j))}{PS(\boldsymbol{Z}_j)}+\hat{\mu}_1(\boldsymbol{Z}_j)\right]-\frac{1}{n}\sum\left[\frac{(1-X_j)(Y_j-\hat{\mu}_0(\boldsymbol{Z}_j))}{1-PS(\boldsymbol{Z}_j)}+\hat{\mu}_0(\boldsymbol{Z}_j)\right] \quad (1)$$

where $\hat{\mu}_i(\boldsymbol{Z}_j)=\hat{E}\left(Y\middle|X=i,\boldsymbol{Z}_j\right)$ is the predicted outcome by regressing Y on $\boldsymbol{Z}$, given $X=i$.

EB is a preprocessing technique proposed by [9] to achieve covariate balance with a binary exposure. Similar to IPTW, EB involves a reweighting scheme that is applied to the sample units, but the weight is formulated by directly incorporating the constraint of moments of a given covariate Z.

There is a lack of consensus on how the preprocessing step on covariate selection is best conducted for propensity and entropy approaches. PS is usually estimated by logistic regression as the prediction of $\mathrm{P}(X=1|\boldsymbol{Z})$ with a certain variable selection strategy, such as the stepwise procedure, if p is relatively larger compared to n, although others suggest using a machine learning method (e.g., the gradient boosting model by [10]. On the variable selection of $\mathbf{Z}$ to estimate PS, some studies have suggested using as many available covariates as possible [5,9]. [12] and [13] suggested using all covariates that can increase the precision for predicting X. However, an over-fitted estimate for X may result in a collinearity issue [12]. [15] indicated that using the maximum number of covariates to estimate PS may result in over-fitting, which may lead to the inflation of estimate variance. To circumvent this problem, some studies have proposed methods that use only important confounders rather than covariates for inferencing an $X \rightarrow Y$ causal model [14-15].

[14] conducted an experimental simulation using three independent covariates, where one was a confounder that associated with both X and Y, one associated with Y but not X, and one associated with X but not Y. By assuming the above relation was known, the authors examined the empirical bias, variance, and mean squared error (MSE) for various covariate combinations in the PS-estimation step, followed by a PS-balanced inferencing step, such as spline regression or stratification. Their results indicated that including the first two covariates to estimate PS gave the smallest bias and MSE compared to other covariate combinations. In this study, we extended the experiment in the sense of using more complicated covariate structures and more recently developed balancing approaches. Moreover, instead of including the Y-but-not-X-associated covariates into the estimation of the balancing score ($e(\boldsymbol{Z})$), we propose a covariate partition-deletion (*cpd*) preprocessing that uses the confounders only in the estimation step, while the Y-but-not-X-associated covariates are adjusted in the next step of inferencing the $X \rightarrow Y$ causality.

When comparing the performance among covariate-balancing procedures, in addition to assessing the bias/MSE and power, it is also important to examine the influence of over-dispersion in the estimation that results in high type I error inflation [16]. Some covariate selections have small bias and variance, but severe over-dispersion, that is, the true (or empirical) standard error (*se*) is larger than the nominal *se* actually used in testing. In this study, we compare the performance of procedures by assessing the MSE, ratio of nominal *se* to empirical *se*, and empirical power/type I error rate.

PS and EB weighting approaches typically only deal with a single exposure variable. However, it is not uncommon that the exposure interacts with other exposures of interest on the outcome. In that case, ignoring the interaction and applying standard balancing procedures on the single exposure may produce severely biased results. For example, the Georgia Capital Charging and Sentencing Study [19] aimed to investigate whether the race of defendant (X_1) and/or the race of victim (X_2) affected the length/severity of the sentence

(Y). In such a study, not only are there two exposures of interest but also significant interaction on the outcome.

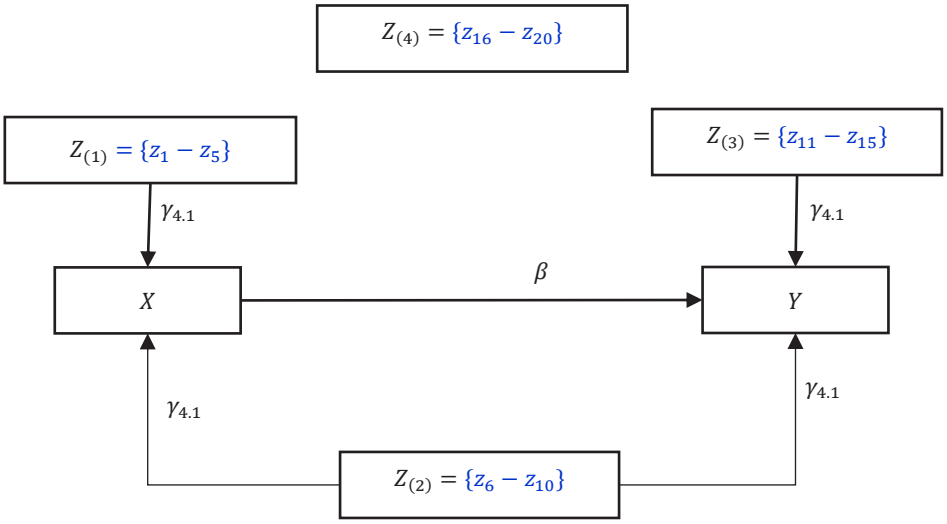
In this paper, we propose a preprocessing step for covariate-balancing for single exposure as well as multiple exposure variables with interaction effects on the outcome. We evaluate performance by simulation and real data analysis.

2 Covariate partition and deletion (*cpd*)

For response Y and exposure X, we first partition the covariates $\boldsymbol{Z}$ into four subsets as shown in Figure 1(A), where

$\boldsymbol{Z}_{(1)}$: subset of $\boldsymbol{Z}$ elements that are significantly associated with X but not Y;
$\boldsymbol{Z}_{(2)}$: subset of $\boldsymbol{Z}$ elements that are significantly associated with both X and Y;
$\boldsymbol{Z}_{(3)}$: subset of $\boldsymbol{Z}$ elements that are significantly associated with Y but not X; and
$\boldsymbol{Z}_{(4)}$: subset of $\boldsymbol{Z}$ elements that are not directly associated with either X or Y.

(A) Single exposure variable X.



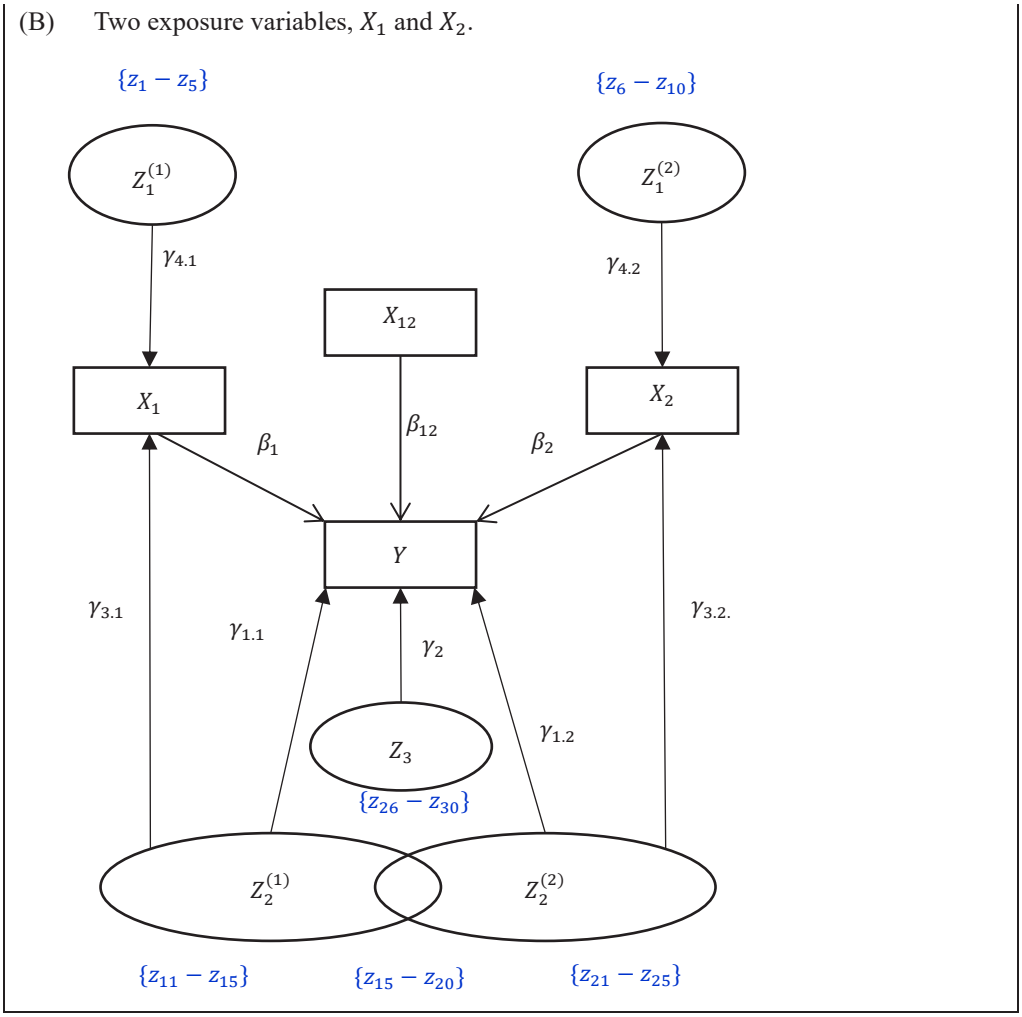


Fig. 1. Settings of covariates **Z** and the association with exposure X and outcome Y.

In this work, the partition is performed by regression selection procedures of Y on $\boldsymbol{Z}$ and X on $\boldsymbol{Z}$, respectively, given a significant level at p-value < 0.15. Then, we define $\boldsymbol{Z}_{(1)}$ to $\boldsymbol{Z}_{(4)}$ as above. After the partition, we suggest the following procedure (refer to as *cpd-2*):

- Delete sets $\boldsymbol{Z}_{(1)}$ and $\boldsymbol{Z}_{(4)}$
- Use $\boldsymbol{Z}_{(2)}$ in the estimation step
- Adjust for $\boldsymbol{Z}_{(3)}$ in inferencing step.

Note that the procedure suggested by [14] (refer to as *cpd-1*) differs from *cpd-2* as $\boldsymbol{Z}_{(3)}$ is used in the estimation step rather than in the inferencing step.

2.1 Simulation setting for single-exposure scenario

We used the following parameter settings:

(1) Covariates $\boldsymbol{Z} = (Z_1, \ldots, Z_{20}) \sim$ Multivariate Normal$(\boldsymbol{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a 20 × 20 matrix with all diagonals equal to 1 and all off-diagonals equal to ρ.

(2) Exposure $X \sim$ Bernoulli (0,1) with $\mathrm{P}(\mathrm{X} = 1) = \exp(\mathrm{f}(\boldsymbol{Z}))/(1 + \exp(\mathrm{f}(\boldsymbol{Z}))$, where

$f(z) = \gamma_1(Z_1 + \cdots + Z_5) + \gamma_2(Z_6 + \cdots + Z_{10})$.

(3) Response $Y = \beta X + \gamma_3(Z_6 + \cdots + Z_{10}) + \gamma_4(Z_{11} + \cdots + Z_{15}) + \varepsilon$, where $\varepsilon \sim N(0,1)$ independent of $\boldsymbol{Z}$.

Then, $\boldsymbol{Z}_{(1)} = \{Z_1, \ldots, Z_5\}$, $\boldsymbol{Z}_{(2)} = \{Z_6, \ldots, Z_{10}\}, \boldsymbol{Z}_{(3)} = \{Z_{11}, \ldots, Z_{15}\}, \boldsymbol{Z}_{(4)} = \{Z_{16}, \ldots, Z_{20}\}$ are subsets of $\boldsymbol{Z}$ that follow the association map in Figure 1(A). Various combinations of parameters $(\beta, \gamma_1, \gamma_2, \gamma_3, \gamma_4)$ are simulated under two covariance designs: independent covariates with $\rho = 0$ and dependent covariates with $\rho \sim \text{Unif}(-0.1, 0.1)$.

Using the simulated data described above, we compared five commonly used balancing procedures: regression-based PS-adjustment, IPTW, EB, a recently developed generalized boosted modeling for PS with the *twang* package in R, and the DR estimation [5,8]. For each procedure, the following three covariate-selection strategies were applied.

1. Strategy (*all-c*): estimate PS or weight by stepwise regression of *X* on all covariate $\boldsymbol{Z}$ in the estimation step, and then use PS or weight for balancing in the inferencing step.
2. Strategy (*cpd-1*): partition $\boldsymbol{Z}$ into $\boldsymbol{Z}_{(1)} \sim \boldsymbol{Z}_{(4)}$ to estimate PS by fitting X on $\boldsymbol{Z}_{(2)} + \boldsymbol{Z}_{(3)}$ in the estimation step, and then regressing Y on X + PS or by weighted regression of Y on X in the inferencing step.
3. Strategy (*cpd-2*): estimate PS by fitting X on $\boldsymbol{Z}_{(2)}$ in the estimation step and regress *Y* on *X* + *PS* + $\boldsymbol{Z}_{(3)}$ or carry out weighted regression of Y on X +$\boldsymbol{Z}_{(3)}$ in the inferencing step.

For example, in IPTW(*all-c*), PS($\boldsymbol{Z}$) is estimated by fitting *X* on all $\boldsymbol{Z}$ covariates, and then inferencing the association by the weighted regression of *Y* on *X* with weights $\frac{1}{PS(Z)}$ and $\frac{1}{1-PS(Z)}$ for the exposure and control groups, respectively. IPTW(*cpd-1*) is the same as IPTW(*all-c*), except that PS = PS($\boldsymbol{Z}_{(2)}$) is estimated by fitting X on the confounders $\boldsymbol{Z}_{(2)}$ only, and IPTW(*cpd-2*) is the same as IPTW(*cpd-1*) except that in the inferencing step, *Y* is regressed on *X* with $\boldsymbol{Z}_{(3)}$ in the weighted regression with weights $\frac{1}{PS(\boldsymbol{Z}_{(2)})}$ and $\frac{1}{1-PS(\boldsymbol{Z}_{(2)})}$. The above combinations are compared with (1) Un-adjusted: regressing Y on X only, (2) Reference 1: regressing Y on X + $\boldsymbol{Z}_{(2)}$ (assuming $\boldsymbol{Z}_{(2)}$ is known), and (3) Reference 2: regressing Y on X + $\boldsymbol{Z}_{(2)} + \boldsymbol{Z}_{(3)}$ (assuming both $\boldsymbol{Z}_{(2)}$ and $\boldsymbol{Z}_{(3)}$ are known).

3 Simulation settings for two-exposure scenario

(1) Covariates $\boldsymbol{Z} = (Z_1, \ldots, Z_{35}) \sim \text{Multivariate Normal}(\boldsymbol{0}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}$ is a 25 × 25 matrix with diagonals equal to 1 and off-diagonals equal to ρ.

(2) Exposure $X_i \sim$ Bernoulli $(0, P(X_i = 1|\boldsymbol{Z}))$ with $P(X_i = 1|\boldsymbol{Z}) = \exp(f_i(\boldsymbol{Z}))/(1 + \exp(f_i(\boldsymbol{Z}))$, where $f_i(Z) = \gamma_{3.i} {Z_2}^{(i)} + \gamma_{4.i} {Z_1}^{(i)}$, *i*=1,2.

(3) Response $Y = \beta_1 X_1 + \beta_2 X_2 + \beta_{12} X_1 X_2 + \gamma_{1.1} {Z_2}^{(1)} + \gamma_{1.2} {Z_2}^{(2)} + \gamma_2 Z_3 + \varepsilon$, where $\varepsilon \sim N(0,1)$ are independent noises, and ${\boldsymbol{Z}_1}^{(1)}: \{Z_1 \ldots Z_5\}$, ${\boldsymbol{Z}_1}^{(2)}: \{Z_6 \ldots Z_{10}\}$, ${\boldsymbol{Z}_2}^{(1)}: \{Z_{11} \ldots Z_{15}, Z_{16} \ldots Z_{20}\}$, ${\boldsymbol{Z}_2}^{(2)}: \{Z_{15} \ldots Z_{20}, Z_{21} \ldots Z_{25}\}$, $\boldsymbol{Z}_3: \{Z_{26} \ldots Z_{30}\}$, and $\boldsymbol{Z}_4: \{Z_{31} \ldots Z_{35}\}$ follow the association map in Figure 1(B). Various combinations of parameters $(\beta_1, \beta_2, \beta_{12}, \gamma_{1.1}, \gamma_{1.2}, \gamma_2, \gamma_{3.1}, \gamma_{3.2}, \gamma_4)$ are simulated under two covariance designs: independent covariates with $\rho = 0$ and dependent covariates with $\rho \sim \text{Unif}(-0.1, 0.1)$.

3.1 Assessment of performance

For each of the N = 1000 replications in the simulation, we recorded the estimate of effect size $\hat{\beta}$ and its nominal standard error ($nom.se$) produced by software, and computed the empirical standard deviation from the N replication, which is regarded as the estimate of true standard error ($emp.se$) for $\hat{\beta}$. Note that $nom.se > emp.se$ results in smaller Wald's statistics and, consequently, a conservative inference, whereas $nom.se < emp.se$ results in an over-optimistic inference, also known as over-dispersion. When true $\beta = 0$, the consequence of over-dispersion is the inflation of type I error.

For a given level of significance $\alpha = 0.05$, the empirical power and type I error are the proportions of rejecting the null hypothesis for the given $\beta \neq 0$ and $\beta = 0$, respectively. The MSEs $\sum_1^N(\hat{\beta} - \beta)^2/N$, along with the over-dispersion ratio $nom.se/emp.se$ and empirical power or type I error rate, are calculated to compare performance among the 15 procedures.

3.2 Procedure for two-exposure variables with interaction effect.

Consider the case with two binary exposures of interest, X_1 and X_2, with possible interaction (denoting X_1*X_2 by X_{12}). We suggest the following procedure for the interaction effect using the DR estimation package *drgee* @R by [20].

(1) Treat X_1 as the main exposure and partition $\boldsymbol{Z}$ into $\boldsymbol{Z}_{(1)} - \boldsymbol{Z}_{(4)}$ according to the association with X_1.
(2) (ii) In *drgee*, estimate X_1 with the selected $\boldsymbol{Z}$ in the exposure formula, use X_2 in the interaction formula, and selected $\boldsymbol{Z}$ elements in the outcome formula to obtain the inferences on X_1 and X_{12}. Note that if *all-c* strategy is used, the selected $\boldsymbol{Z}$ elements will be all covariates $\boldsymbol{Z}$, whereas if *cpd-2* is used, this will be $\boldsymbol{Z}_{(2)}$ in the exposure formula and $Z_{(3)}$ in the outcome formula.
(3) (iii) Reverse the positions of X_1 and X_2 and repeat steps (i) and (ii) to obtain the inferences on X_2 and another X_{12}.

With two inferences on interaction X_{12}, a conservative way to conclude the significance if both X_{12} are significant, and an aggressive way to conclude significance if at least one is significant.

4 Results

4.1 Comparison of covariate selection strategies for single-exposure scenarios

Under $(\gamma_1, \gamma_2, \gamma_3, \gamma_4) = (0.6, 0.2, 0.2, 06)$ and $\rho \sim \text{Unif}(-0.1, 0.1)$, supplementary Figures (S.Fig 1-3) display the results for β = 0, 0.2, and 0.4, respectively. For $\beta = 0$, S.Fig 21shows the comparison of empirical type I error rate, MSE, and over-dispersion ratio of ($nom.se/emp.se$) across three covariate-selection strategies and five balancing procedures. S.Fig 2 and 3, with $\beta = 0.2$ and 0.4, show the same statistics, except with power instead of type I error. Other parameter settings $(\gamma_1, \gamma_2, \gamma_3, \gamma_4)$ were also tried, but all showed similar tendencies as in S.Fig 1-3.

In S.Fig 1, the order of MSE magnitude among the three strategies was (*all-c*) > (*cpd-1*) – (*cpd-2*) across all five balancing procedures. When (*all-c*) was applied, the MSE magnitude across all β was *twang* > IPTW > EB ~ PS-reg ~ DR.

All the procedures had reasonable type I error rates except for twang(*all-c*) and IPTW(*all-c*). IPTW(*all-c*) also had the lowest ratio of $nom.se/emp.se$ ~ 0.83–0.85, which results in

a 1.2-fold over-dispersion of the T-statistics. In S.Fig 2 and 3, *cpd-2* has the best power and the least MSE across all five balancing procedures.

Figure 2 summarizes the power for different procedures. For DR, the order of power among the three strategies was (*cpd-2*) – (*cpd-1*) > (*all-c*) , while that for the other four procedures was (*cpd-1*) < (*all-c*) < (*cpd-2*) for both $\beta = 0.2$ and $\beta = 0.4$.

When (*cpd-2*) was applied, the power at $\beta = 0.2$ for DR, EB, IPTW, and *twang* were all around 0.4~0.5, slightly more powerful than PS-reg at 0.34. At $\beta = 0.4$, DR has slightly better power (0.96) than the other procedures (0.91–0.93). All results suggest that (*cpd-2*) has the best power and reasonable type I error rate.

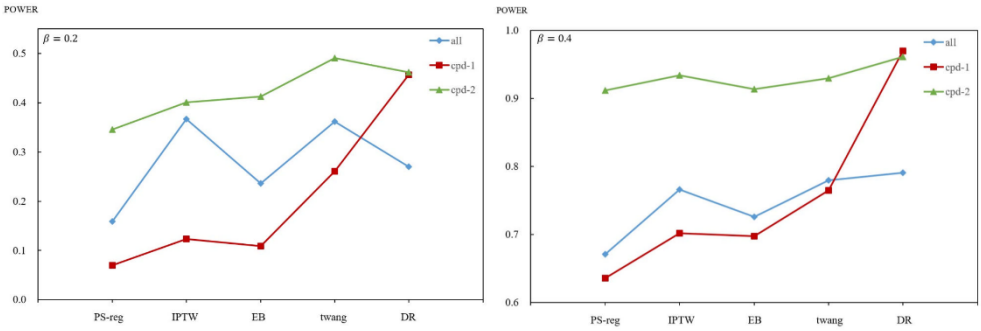


Fig. 2. Power under three covariate selection strategies for five balancing procedures.

Link for Supplementary Figures:
https://www.dropbox.com/scl/fi/p3sjxyogjsfa2nm2lm42f/Supplemetary_Fig.pdf?rlkey=nvh9tfsnzy2me7ny2ir8abti3&dl=0

4.2 Comparison of covariate-selection strategies for two-exposure scenario

Under various experimental settings of $(\beta_1, \beta_2, \beta_{12}, \gamma_{1.1}, \gamma_{1.2}, \gamma_2, \gamma_{3.1}, \gamma_{3.2}, \gamma_4)$ and ρ, we could not find any dominant strategy due to the complexity of the covariate structures; therefore, we omit the results. In particular, when there are covariates associated with both exposure variables, the partition strategies may not be as successful as in single exposure scenarios. In general, although strategy *all-c* still results in higher over-dispersion ratios (nominal.se/empirical.se) on the estimate of the interaction effect, ranging from 1.1 to 1.06 in the simulation trials, its power is relatively more robust than *cpd-1* or *cpd-2*. Therefore, we suggest using the *all-c* strategy for two-exposure scenarios, but being cautious about the associated collinearity impacts on the inflation of the type I-error rate and estimated standard error.

4.3 Georgia Capital Charging and Sentencing Study

Charging and sentencing data were obtained from legal records in Georgia for cases tried between March 1973 and December 1979. The purpose of the study was to investigate charging and sentencing in non-negligent homicide cases. The data were extensively analyzed by [19]. The data analyzed here are 1077 murder and voluntary manslaughter cases with 275 variables including information on the victims, defendants and how the crimes were committed, with items such as number of prior convictions, number of prior arrests, and so on. Among the 1077 cases in the data, 512 were sentenced to a limited term of imprisonment ($Y = 0$), and 565 otherwise ($Y = 0$), including 423 of life imprisonment and 142 of death penalty. The two exposures of interest in the original study are race of defendant (BLD: 1 =

black, 0 = o/w) and race of victim (WHV: 1 = white, 0 = o/w). We applied a principal component analysis to the rest remaining 272 variables, which were considered as background variables in this analysis, and found that the first 50 factors together account for 90% of the variance. These 50 factors were used as covariates (***Z***) in the following analyses. Table 1 shows their DR estimation under two models (with/without interaction) using two covariate selection strategies (*all* vs *cpd-2*). Since the *drgee* package @R only allow one main exposure at time, we fitted two models that one with BLD as exposure and WHV as the covariate that mediates BLD, and another with the reverse positions. Therefore, there are two estimates of interaction WHV*BLD for each covariate selection strategy in Table 1.

In the models without interaction, WHV was significant under both selection strategies (p-value < 0.001), but when interaction was modelled, WHV became insignificant and the interaction BLD*WHV showed potential impact on the outcome. Under the *cpd-2* covariate selection, the interaction BLD*WHV was highly significant with p-value 0.0005 and estimated coefficient 2.347 using BLD as exposure and WHV as covariate, and with p-value 0.0067 and estimated coefficient 1.571 using WHV as exposure and BLD as covariate. On the other hand, although the *all-c* strategy had similar estimated coefficients to those under *cpd-2*, its standard errors seemed inflated compared to other strategies.

Table 1. Doubly robust estimation of two exposure effects on the severity of sentence, under two models (with/without interaction) with two covariate selection strategies (*all* vs *cpd-2*).

Model		Effect	Est.coef	Std.err	P
No-interaction	(*all-c*)	BLD[a]	-0.379	0.341	0.2660
		WHV[a]	1.177	0.303	**<0.001***
	(*cdp-2*)	BLD*	0.400	0.271	0.1400
		WHV*	1.979	0.261	**<0.001***
With interaction	(*all-c*)	BLD	-1.854	0.597	**0.002***
		WHV*BLD	2.968	2.195[b]	0.176
		WHV	-0.270	0.569	0.835
		WHV*BLD	2.788	1.328[b]	**0.036**
	(*cdp-2*)	BLD	-1.739	0.451	**0.0001***
		WHV* BLD	2.347	0.673	**0.0005***
		WHV	-0.328	0.429	0.445
		WHV* BLD	1.571	0.579	**0.0067***

a BLD: race of defendant (1 = black, 0 = o/w), WHV: race of victim (1 = white, 0 = o/w).
b extra large se of all-c strategy comparing to cpd-2.
* significant at level 0.05

5 Discussion

In this paper, we investigated several covariate-preprocessing strategies for single-exposure and two-exposure-variables scenarios under commonly used balancing procedures such as DR estimation, PS-regression, EB weighting, IPTW and the gradient boosting model. These procedures comprise two steps: modeling the exposure to estimate balancing score (weight) and modeling the outcome to infer the causal effect. Based on simulation and real-data analysis of performance, we suggest using the *all-covariate* strategy for a two-exposure scenario, especially where the interaction effect exists. However, caution needs to be applied in relation to the inflation of type I-error rate and estimated standard error due to collinearity when using all covariates. By contrast, for a single-exposure scenario, we suggest using a covariate-selection strategy such as *cpd-2* to reduce over-dispersion on the estimate of the exposure-outcome effect. Note that the main difference between *all-c* and *cpd* is that the X-but-not-Y-associated covariates are used in the estimation step for the *all-c* strategy but not for *cpd*. The difference between *cpd-1* and *cpd-2* is that Y-but-not-X-associated covariates are used in the inferencing step in *cpd-2* but not in *cpd-1*. Therefore, our results imply the following insights:

(A) it is critical that the X-but-not-Y-associated covariates should be in neither the PS-estimation step (1) nor the inferencing step (2) for all balancing procedures,
(B) including the Y-but-not-X-associated covariates in step (2) to model the outcome can improve the power for most of the procedures, and
(C) DR with *cpd-2* strategy has the best performance among all procedures.

Including the X-but-not-Y-associated covariates in the PS-estimation step may result in over-dispersion for the causal inference of IPTW and gradient boosting model (*twang*), resulting in type-I-error inflation. By contrast, neglecting Y-but-not-X-associated covariates in the inferencing step may result in under-dispersion for PS-regression, EB, IPTW and the gradient boosting model, resulting in lower power.

The DR estimation is known for its robustness as the regressions in the PS-estimation step and outcome-inferencing step may alleviate model misspecifications between the two. However, our results show that DR can only remedy the neglect of some confounders by one of the steps (i.e., *cpd-1* and *cpd-2* performed equally well), but cannot remedy the inclusion of the X-but-not-Y-associated covariates. When X-but-not-Y-associated covariates were included in modeling PS-estimation (e.g., *all-c*), the power for DR decreased from 0.46 to 0.27 for $\beta = 0.2$ and from 0.97 to 0.79 for $\beta = 0.4$.

One of the limitations of this study is that we only considered linear relations for the covariate-outcome-exposure association. In addition, in the cases with high correlation among covariates, the partition of covariates may not be accurate. One way to mitigate this problem is to transform the correlated covariates into orthogonal covariates by multivariate techniques, such as principal component analysis.

References

1. S. Tübbicke, J. Econom. Methods, **11**(1), 71-89 (2022) https://doi.org/10.1515/jem-2021-0002
2. P. R. Rosenbaum, D. B. Rubin, Biometrika, **70**(1), 41-55 (1983).
3. A. Markoulidakis, K. Taiyari, P. Holmans, P. Pallmann, M. Busse, M. D. Godley, B. A. Griffin, Health Serv. Outcomes Res. Methodol., **23**(2), 115–148 (2023) https://doi.org/10.1007/s10742-022-00280-0
4. K. Narita, J. D. Tena, C. Detotto, Leadersh. Q., **34**(3), 1-16 (2023)

5. D. O. Scharfstein, A. Rotnitzky, J. M. Robins, J. Am. Stat. Assoc. **94**(448), 1121–1146 (1999)
6. H. Bang, J. M. Robins, Biometrics, **61**(4), 962–973 (2005)
7. J. Robins, M. Sued, Q. Lei-Gomez, A. Rotnizky, Stat. Sci. **22**(4), 544–559 (2007)
8. M. J. Funk, D. Westreich, C. Wiesen, T. Stürmer, M. A. Brookhart, M. Davidian, Am. J. Epidemiol., **173**(7), 761-7 (2011) doi: 10.1093/aje/kwq439. Epub 2011 Mar 8. PMID: 21385832; PMCID: PMC3070495.
9. J. Hainmueller, Political Anal., **20**(1), 25-46 (2012)
10. D. F. McCaffrey, B. A. Griffin, D. A. Almirall, M. E. Slaughter, R. Ramchand, L. F. Burgette, Stat. Med. **32**(19), 3388-3414 (2013) doi:10.1002/sim.5753
11. G. Ridgeway, D. F. McCaffrey, A. R. Morral, M. Cefalu, L. F. Burgette, J. D. Pane, B. A. Griffin, Toolkit for Weighting and Analysis of Nonequivalent Groups: A Tutorial for the R TWANG Package. Santa Monica, CA: RAND Corporation (2022) https://www.rand.org/pubs/tools/TLA570-5.html.
12. S. Weitzen, K. L. Lapane, A. Y. Toledano, A. L. Hume, V. Mor, Phamacoepidemiol. Drug Saf., **13**, 841–53 (2004)
13. T. Stürmer, M. Joshi, R.J. Glynn, J. Avorn, K. J. Rothman, K. J., & Schneeweiss, S., J. Clin. Epidemiol., **59**(5), 437–447 (2006) https://doi.org/10.1016/j.jclinepi.2005.07.004
14. M. A. Brookhart, S. Schneeweiss, K. J. Rothman, R. J. Glynn, J. Avorn, T. Stürmer, Am. J. Epidemiol., **163**(12), 1149–1156 (2006). https://doi.org/10.1093/aje/kwj149
15. T. Schuster, W. K. Lowe, R. W. Platt., J. Clin. Epidemiol., **80**, 97-106 (2016) doi: 10.1016/j.jclinepi.2016.05.017. Epub 2016 Sep 14. PMID: 27498378; PMCID: PMC5756087.
16. S. Schneeweiss, J. A. Rassen, R. J. Glynn, J. Avorn, H. Mogun, M. A. Brookhart, Epidemiology **20**(4), 512-22 (2009). doi:10.1097/EDE.0b013e3181a663cc
17. J. F. Yang, M. Webster-Clark, J. L. Lund, R. S. Sandler, E. S. Dellon, T. Stürmer, Gastrointest. Endosc., **90**(3), 360-369, (2019) doi:10.1016/j.gie.2019.04.236
18. I. B. Lian, Comput. Stat. Data Anal. **43**(2), 197-214 (2003) doi:10.1016/S0167-9473(02)00223-2
19. D. Baldus, G. Woodworth, C. Pulaski, Equal Justice and the Death penalty (1990)
20. J. Zetterqvist, A. Sj"olander, Epidemiol. Methods, 4(1), 69–86 (2015).